Visual attention using spiking neural maps

Roberto A. Vazquez, Bernard Girau, Jean-Charles Quinton

Visual attention using spiking neural maps

Roberto A. Vazquez, Bernard Girau and Jean-Charles Quinton

Abstract— **Visual attention is a mechanism that biological systems have developed to reduce the large amount of visual information in order to efficiently perform tasks such as learning, recognition, tracking, etc. In this paper, we describe a simple spiking neural network model that is able to detect, focus on and track a stimulus even in the presence of noise or distracters. Instead of using a regular rate-coding neuron model based on the continuum neural field theory (CNFT), we propose to use a time-based code by means of a network composed of leaky integrate-and-fire (LIF) neurons. The proposal is experimentally compared against the usual CNFT-based model.**

I. Introduction

Attention enables us to dynamically select and enhance the processing of the most relevant stimuli and events at each moment [1]. Visual attention is a powerful mechanism that enables perception to focus on a small subset ("where to look") of the information picked up by our eyes [2], based on both bottom-up and top-down cues [3].

The control of focal visual attention involves an intricate network of brain areas, spanning from the primary visual cortex to the prefrontal cortex. Selecting where to attend next is primarily controlled by the dorsal visual processing stream (or "where/how" stream) which comprises cortical areas in the posterior parietal cortex, whereas the ventral visual processing stream (or "what" stream), comprising cortical areas in the inferotemporal cortex, is primarily concerned with localized object recognition [4].

Several computational theories and models have been developed for how attention may be attracted towards a particular object in the scene rather than another [5], [6], [7]. Recently, the authors in [8] have demonstrated that bottom-up (i.e. stimulus driven) attention may be seen as an emergent property of a neural population using the Continuum Neural Field Theory. From a pool of neurons spread over two maps, one input map feeding a focus map, a bubble of activity emerges within the focus map at the precise location of a stimulus presented within the input map. Furthermore, when noise or distracters are added, the bubble of activity stays focused on the chosen stimulus and then, between several simultaneously possible objects, the model is able to maintain visual attention onto the one stimulus it first focused.

The goal of this paper is to show that similar attentional properties (detection, focus and tracking of a stimulus) can be obtained by a population of spiking neurons, with an improved robustness in the presence of noise or distracters. The visual attention model we propose is based on a simple spiking neural model, the so-called leaky integrate-and-fire (LIF) neurons. This model explicitly handles temporal events (spikes) that are usually "hidden" in the rate-coding neuron models used by the continuum neural field theory. This proposal is experimentally compared against the CNFT-based model corroborating its robustness to noise and distracters while achieving the main visual functionalities depicted in [8]: competitive behavior, tracking and target switching.

Section II describes the motivation for using spiking neurons within neural models of visual attention, while introducing both CNFT-based and LIF-based models. Our model is precisely defined in section III. Its properties are experimentally validated in section IV.

Roberto Vazquez is with the Intelligent Systems Group, Faculty of Engineering, La Salle University, Benjamín Franklin 47, Col. Condesa, 06140, Mexico City, Mexico (email: ravem@lasallistas.org.mx) - Bernard Girau and Jean-Charles Quinton are with the LORIA/Cortex project - Campus Scientifique, B.P. 239, 54506 Vandoeuvre-lès-Nancy Cedex, France (email: bernard.girau@loria.fr). This work was supported by the INRIA CorTexMex Research Associate Team.

II. Motivation and related work

Neural fields, and more specifically CNFT-based neural fields, have proved very powerful to build neural models able to perform more or less complex visual tasks such as visual attention [8], scene exploration [9], overt attention [10] or motion discrimination [11]. These models are massively distributed, so that we aim at using them in autonomous embedded systems thanks to their hardware parallel implementation. Preliminary attempts such as in [12] have shown that their dense local interactions are too demanding to define efficient parallel implementations of these models. On the other hand, the study in [13] has shown that simple spiking neurons may be efficiently assembled on a hardware device such as an FPGA (Field Programmable Gate Array) thanks to their reduced communication bandwidth: whatever the level of complexity of the internal computations of each neuron, simple 1-bit messages (spikes) are exchanged between neurons. This has motivated the will to define populations of simple spiking neurons that mimic the properties of the CNFT-based models for visual attention. This is not the first attempt to define spiking neural fields. A related approach may be found in [14]. In order to precisely situate the contribution of our model with respect to [8] and [14], we now give more details about these two visual attention models that are able to perform a tracking task of a target even in the presence of very strong noise or in the presence of a lot of distracters.

A. CNFT visual attention model

The Continuum Neural Field Theory is a kind of dynamic neural field model that implements lateral competition within cortical maps [15], [16]. The CNFT can be reduced to a single differential equation that describes the evolution of

the membrane potential of neurons over cortical maps. This is a continuous approximation of the evolution of large populations of neurons that interact through excitatory and inhibitory connections: close neurons tend to be reciprocally excited, distant neurons inhibit themselves.

In [8] the authors propose a visual attention model based on a discrete version of the CNFT. This model is highly robust and it is able to track a static or moving target in the presence of noise with high intensity or despite a lot of distracters possibly more salient than the target. The main hypothesis about the target stimulus is that it has a spatio-temporal continuity that should be observable by the model.

A neural position is labeled by a vector $\mathbf{x}$, which represents a two-component quantity designing a position on a manifold $\mathbf{M}$ in bijection with $[0,1]^2$. The membrane potential of a neuron at position $\mathbf{x}$ and time t is denoted by $u(\mathbf{x},t)$. The lateral connection weight function $w(\mathbf{x}-\mathbf{x}')$ is a difference of Gaussian function applied to distance $|\mathbf{x}-\mathbf{x}'|$. An afferent connection weight $s(\mathbf{x}_{\mathbf{M}'},\mathbf{x}_{\mathbf{M}})$ applies to the local stimulus received at position $\mathbf{x}_{\mathbf{M}}$ in manifold $\mathbf{M}$ from position $\mathbf{x}_{\mathbf{M}'}$ in manifold $\mathbf{M}'$. The membrane potential $u(\mathbf{x},t)$ satisfies the following equation:

$$\begin{array}{ll}\tau\frac{\partial u(\mathbf{x},t)}{\partial t} = -u(\mathbf{x},t) & +h \\ & +\frac{1}{\alpha}\int_M w_M(\mathbf{x}-\mathbf{x}')f[u(\mathbf{x}',t)]\,d\mathbf{x}' \\ & +\frac{1}{\alpha}\int_{M'} s(\mathbf{x},\mathbf{y})\,I(\mathbf{y},t)\,d\mathbf{y}\end{array} \tag{1}$$

where f represents the mean firing rate as some function of the membrane potential u of the relevant cell, $I(\mathbf{y},t)$ is the external stimulus at position $\mathbf{y}$ and time t in $\mathbf{M}'$, h is the neuron threshold and α is a scaling term.

Lateral connection weights are given by:

$$w_M(\mathbf{x}-\mathbf{x}') = Ae^{\frac{|\mathbf{x}-\mathbf{x}'|^2}{a^2}} - Be^{\frac{|\mathbf{x}-\mathbf{x}'|^2}{b^2}} \tag{2}$$

and afferent connection weights are given by:

$$s(\mathbf{x},\mathbf{y}) = Ce^{\frac{|\mathbf{x}-\mathbf{y}|^2}{c^2}} \tag{3}$$

Furthermore, the activity of a neuron is bounded between 0 and 1: if $u(\mathbf{x},t) > 1$, $u(\mathbf{x},t) = 1$, and if $u(\mathbf{x},t) < 0$, $u(\mathbf{x},t) = 0$.

The model consists of two maps, input and focus, each of them being of size $n \times n$ units. The input map corresponds to an entry that is feeding the focus map, whereas the focus map represents a cortical layer whose units possess localized receptive fields on the surface of the input. Each unit $\mathbf{x}_{ij}$ of the focus map receives its input from the input map using Eq. 3 which corresponds to a localized receptive field. While the input map does not have any lateral interaction or feedback, the units in the focus map are laterally connected using a difference of Gaussians.

B. LIF visual attention model

In [14] the authors propose a visual attention model based on Leaky Integrate and Fire neurons. This work can be seen as a direct transformation of the CNFT-based model of [8] by using spike-based computations. It may also be seen as the attentional part of a more general model of covert attention that includes an additional pre-attentional part [17]. This visual attention model is again able to focus and stay focused even when the stimulus moves. In addition, they show that noisy backgrounds and distracters only have a small influence on the behavior of the model. Indeed, the experimental results are close to those observed with a CNFT model, although the CNFT-based model seems to be more accurate than the LIF-based model. The authors also experimentally validate the fast computation time achieved by their model.

Again, this model consists of two maps, input and focus, each of them being of size $n \times n$ units. Depending on the parameters used to model a neuron, this set of neurons can either integrate the information over a predefined temporal window or act as a synchrony detector (emitting spikes when inputs are condensed in a small period of time). Neurons from the input map behave as integrators and neurons from the focus map as synchrony detectors.

The input map translates an input stimulus into spike trains. Each unit of the input map is connected to each unit of the focus map through a Gaussian mask. A mask is a static weight matrix and it defines a generic projection of a neural map onto another. The weight matrix values of the Gaussian mask can be viewed as the parameters of a Gaussian image filter. On the other hand, this projection could be related to the afferent connections given by Eq. 3.

The focus map is laterally connected using a difference of Gaussians as in [8], see Eq. 2, which mutually excites adjacent neighbors and inhibits distant ones. This lateral connectivity alone is not sufficient to maintain a self-sustained activity. For this purpose, it needs the spikes from the input map to have an ongoing activity.

III. Proposed model

Instead of using rate-coded neurons as in [8], the model we propose uses leaky integrate-and-fire neurons as in [14].

A. Definition

Our model mostly consists of a 2D neural map of size $n \times n$ LIF neurons that emit their spikes at discrete times. The membrane potential V_i of neuron i is given by the following differential equation:

$$\tau\frac{\partial V_i}{\partial t} = -g_{leak}(V_i(t) - E_{leak}) + \gamma \cdot I_i(t) \tag{4}$$

where g_{leak} and E_{leak} are the conductance and the reversal potential of the voltage-independent leak current, and τ is the membrane time constant. $I_i(t)$ is the input stimulus given at time t, and γ is a scaling term. Each neuron is associated with a pixel, i.e. the pixel luminance determines the input term $I_i(t)$ of the corresponding neuron. This differential equation is approximated through a simple Euler scheme (see [18] to assert that simple discrete-time generalized LIF neurons have the same expressive power as the underlying continuous models).

Whenever the membrane potential V reaches the threshold θ, a spike is fired, and V is instantaneously reset to rest.

Finally, the membrane potential V_i of neuron i takes into account the influence of incoming spikes by means of the following equation:

$$V_i(t+dt) = V_i(t) + \tau \frac{\partial V_i}{\partial t} dt + \frac{I_i^{syn}(t)}{\gamma} \tag{5}$$

where $I_i^{syn}(t)$ is the synaptic current due to the action of other neurons of the network describing the influence of incoming spikes on the membrane potential. It is given by the following equation:

$$I_i^{syn}(t) = \sum_j w_{ij} S_j(t) \tag{6}$$

where w_{ij} is a weight matrix given by a difference of Gaussians (see Eq. 2) centered at neuron i, which excites adjacent neighbors and inhibits distant ones. Finally, the instantaneous spike $S_j(t)$ is given by:

$$S_j(t) = \begin{cases} 1 & V_j(t) \geq \theta \\ 0 & V_j(t) < \theta \end{cases} \tag{7}$$

B. Model positioning

Although this model uses the same kind of LIF neurons as in the model proposed in [14], there are three main differences that impact the computational cost of the model and its behavior.

First of all, our proposal could be seen as a reduced version of the model in [14], since the focus neural map is the only one composed by LIF neurons. The input map simply sends the stimulus information to the focus map. Therefore the input stimulus is not transformed into spike trains as in [14], it directly feeds the focus map. This simplification not only induces a lower computational cost[1], above all it has a strong influence on the future ability of the model to be mapped onto hardware parallel devices, since it avoids the complex connection topology required by two neural maps interconnected in a retinotopical way with multiple overlapping receptive fields.

Another difference is related to the use of filters. In [14] the output of the input map is filtered by means of a Gaussian filter and then it is sent to the focus map. Our model does not apply any Gaussian filter, the receptive fields in the input map being directly received by the focus map without any type of weighted afferent connections. The lateral weights of the focus map appear as sufficient to spatially smooth out the activity of the neurons over small receptive fields.

A third difference is the γ scaling term that appears both in the evolution of the membrane potential with respect to the received stimulus and in the influence of the incoming spikes. This parameter enables us to easily balance both kinds of influences. The experimental results in section IV includes a detailed study of the impact of such a balance on both robustness and behavior of the model.

Finally, we not only propose a LIF-based model that exhibits similar properties as the initial CNFT-based model, we perform a large experimental study so as to quantitatively compare the degree of robustness of both models.

[1]The architecture of [14] is based on a one-to-one correspondance between both maps of LIF neurons. The computational cost mainly lies in the constant update of neuron potentials in both maps. Therefore a rough approximation of the computational cost of [14] is twice the computational cost of our model.

IV. Experimental results

Before studying the robustness of the model in the presence of perturbations, it is necessary to validate its ability to provide the attentional properties that are typical of the usual CNFT-based model.

A. Experimental setup

In order to test the proposed model, we use an experimental setup similar to [8] and [14]: a stimulus follows a circular path on a $n \times n$ pixels input image with either noise or distracters in the background. The stimulus is a Gaussian patch whose center is localized at (x_c, y_c) of width W and intensity I given by:

$$S(x, y, r, \theta, W, I) = I e^{\frac{-(x-x_c)^2}{W^2}} e^{\frac{-(y-y_c)^2}{W^2}} \tag{8}$$

where $x_c = r \sin\theta$ and $y_c = r \cos\theta$.

On the one hand, distracters are exact copies of the Gaussian patch, but they lack spatio-temporal continuity. On the other hand, the added noise is assumed to be independently drawn from a Gaussian distribution with different variance levels σ and mean zero.

Parameters for generating the synaptic weight connections are defined as: $\alpha = n/2$, $A = 25/\alpha$, $a = 5/n$, $B = 12.5/\alpha$ and $b = 75/n$. It must be noted that these weights could be tuned to optimize each type of expected behavior as in [19], but in this work, we rather adopt the approach of [8] where the authors study the ability of the model to simultaneously exhibit all properties with the same set of weights that are experimentally chosen.

During the integration step $dt = 0.1$, $\tau = 1$. Based on eq. 4, we choose to increase the influence of the input compared to the lateral connections by setting $\gamma = 10$. This gaining factor helps the neurons in the focus map to rapidly generate a spike. The exact influence of γ is analyzed in IV-C.

As in [8] and [14], the input image only contains the stimulus target until the first spikes appear on the focus map. In order to ensure that the spikes emerge in the focus map, we wait ten computational steps (one second) before the Gaussian patch starts to move, or before adding some distracters or noise. After that, we generate a new image each 10 computational steps.

To validate the accuracy of the proposal, we compute the normalized euclidean distance between the stimulus center and the center of the activity in the focus map. Following the same definition as in [14], the center of activity is defined as the center of all spikes emitted by the focus map during the integration steps of the image presentation. This center of activity is computed as in [8] by means of the following equation:

$$(x_c, y_c) = \left(\frac{\sum_{i,j} \frac{i}{n} a(i,j)}{\sum_{i,j} a(i,j)}, \frac{\sum_{i,j} \frac{j}{n} a(i,j)}{\sum_{i,j} a(i,j)} \right) \quad (9)$$

where $a(i,j)$ is the value at position (i,j).

As explained in this section, our experiments use synthetic inputs, since we aim at comparing the models with the same criteria already used in similar studies ([8], [14], [19]) that clearly quantify the visual attention properties alone. The question of how the models perform on real video sequences is not addressed in this paper. Nevertheless, it must be noted that all these attentional models may be used for real video sequences, adding a pre-attentional processing part (as in [17] or [12]), but in this case the performance of the model strongly depends on the pre-attentional processing and its ability to properly extract moving objects.

B. Behavioral properties

Following the scenarios defined in [19] to characterize the main expected properties of visual attentional models, we may summarize these properties as competition (ability to focus on a specific target when multiple targets appear in the visual stimulus), tracking (ability to track a focused target that moves), switching (ability to focus on a new target when the previous one disappears), and tolerance to perturbations (noise and/or distracters). Since the latter is directly linked to the study of the robustness of the model in IV-C, we consider here the three first attentional properties.

1) Competitive behavior: The proposed model exhibits a competitive behavior with the same set of synaptic weights used throughout this paper. Furthermore, even when targets are too close, the model is able to focus on only one target after a small number of computational steps.

The only constraint that should be satisfied is that for a pair number of targets, they should not be placed regularly and equidistantly from the center of the focus map: this is a well-known side effect of pure synchronous evaluation when interactions are perfectly symmetrical, that results in all bubbles cyclically appearing and disappearing. Any level of noise of asynchronous evaluation avoids this side effect [20].

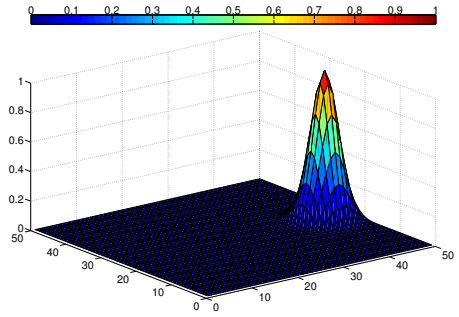
2) Tracking: The model is able to track a focused target that moves in the visual scene. Nonetheless, we found that the velocity at which the targets update their position should be at least two computational steps, with the chosen setup. Again, this is a well-known property of the CNFT-based model: when the target moves too fast, it rapidly goes out of the excitatory range of the emerged bubble, and thus becomes unable to attract it towards the new position of the target. As expected, we also observed that faster tracking can be achieved with an appropriate set of synaptic weights.

3) Switching: Again, the qualitative behavior is preserved: when the initial target disappears, a self-sustained bubble of activity remains a few computational steps at the last location of the bubble, but its amplitude decreases and the system finally focuses on another target. In addition, we observed that the reactivity of the model is strongly linked to the spiking threshold of the neurons: a higher threshold reduces the ability of the bubble to have a long self-sustained activity.

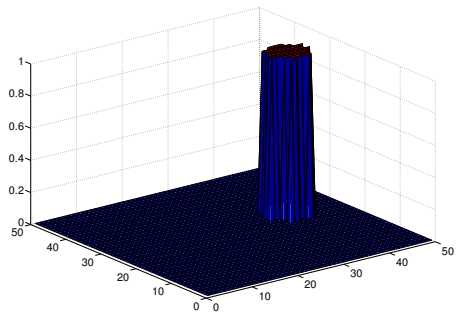
C. Robustness

Three types of experiments were performed to evaluate the robustness of the proposal. The first set of experiments evaluates the robustness of the proposal in the presence of noise. The second set of experiments verifies the robustness of the proposal in the presence of distracters. Finally, in the third set of experiments, we study the effect of changing the gaining factor γ in the accuracy and reactivity of the proposed model. In all experiments, we consider perturbations added to a bell-shaped input such as depicted in figure 1.

In [8] and [14], the authors assume that noise or distracters are added each computational step. It is important to notice that in those experiments the time that noise or distracters remain in the same position might be too short to significantly alter the behavior of the model. In this paper we study the behavior of the model when the noise or distracters are updated at the same rate as the target as well as when distracters and noise are updated at a slower or faster rate than the target, so as to assert that some results are not obtained only thanks to an unfair instability of the perturbations, as it might sometimes be the case in [8] and [14].

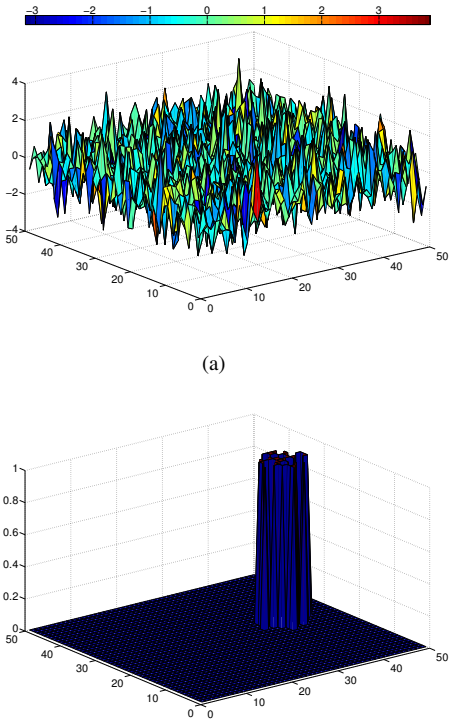


(a)



(b)

Fig. 1. Input is a bell-shaped curve centered around (x_c, y_c) representing an external input. This information is received by the input map which is directly connected to the focus map. (a) Noiseless input. (b) Neurons firing in the focus map as a response to the noiseless input.



(a)

(b)

Fig. 2. Noisy input. The added noise is assumed to be independently drawn from a Gaussian distribution. (a) Noisy input with $\sigma = 1$. (b) Neurons firing in the focus map.

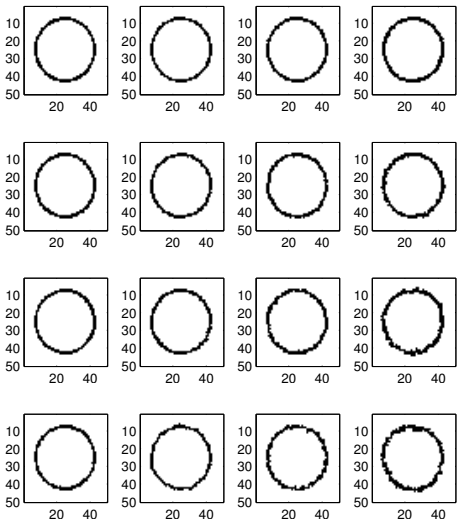


Fig. 3. Trajectories followed by the bubble of activity in the focus map when the target changes its position each ten computational steps along its circular trajectory. Each column presents the trajectory followed by the bubble when the input is perturbed by different levels of Gaussian noise (0.2, 0.4, 0.6 and 0.9). Each row presents the trajectory followed by the bubble when the noise is updated at different rates (1, 5, 10 and 20 computational steps).

1) Robustness of the model in the presence of noise: During the first set of experiments we could observe that the proposed model was able to track the stimulus even when the intensity of the noise was greater than the input stimulus (see Figure 2(a) and Figure 2(b)). The obtained results show that the proposed model is highly robust to the presence of noise (see Figure 4(a)). Despite the high intensity of the noise added to the input, the proposed model stays focused on the input target (see Figure 3).

As was demonstrated in [8], the CNFT visual attention model also provides highly acceptable results when the target is perturbed by some noise (see Figures 4(b)). However, the accuracy of the CNFT model starts to decrease when the intensity of the noise added to the background exceeds the intensity of the target (variance greater than 0.5).

As can be observed from Figures 4(a) and 4(b), the time that noise remains without change modifies the accuracy of the model. If noise changes rapidly, for example every computational step (0.1 second), the accuracy that the model provides is higher than the accuracy achieved when the noise changes slowly. This phenomenon may be easily explained by the inertia of the bubble of activity that is less disturbed by transient perturbations. Nevertheless, the accuracy obtained with the proposed method is still highly acceptable when noise is updated every two seconds. On the contrary, under the same conditions, the accuracy achieved by the CNFT-based model decreases more rapidly when noise shows some inertia.

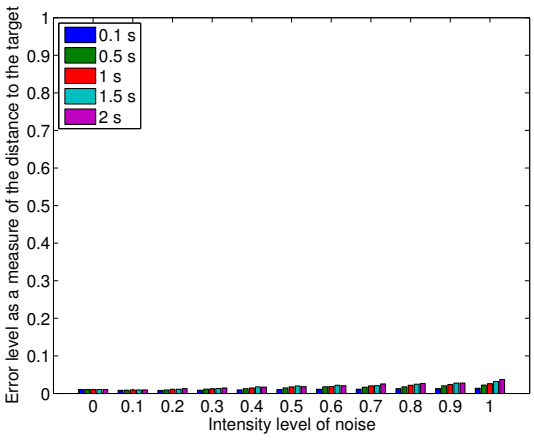
The LIF visual attention model of [14] also provides highly acceptable results when the target is perturbed by noise, though the influence of the noise inertia is not studied. Anyway, our model provides comparable results even when the background is highly noisy, despite its simplicity (one map of LIF neurons).

2) Robustness of the model in the presence of distracters: During the second set of experiments, we observed that the proposed model was also robust in the presence of several distracters (even more salient than the target), see Figure 5(a) and Figure 5(b). The obtained results show that the proposed model is highly robust to the presence of distracters (see Figure 7(a)). Despite the number of distracters added to the input, the proposed model stays focused on the input target (see Figure 6).

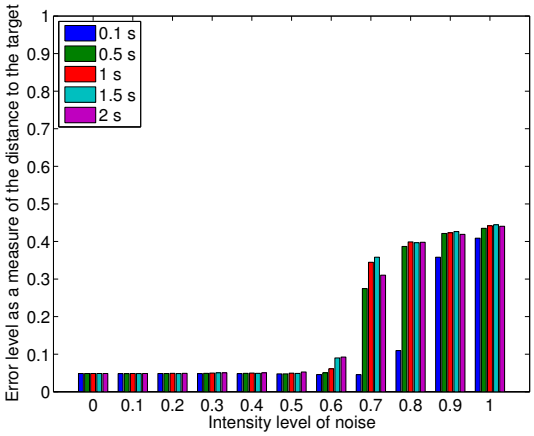
On the other hand, as was stated before, the CNFT visual attention model also provides highly acceptable results when the target is perturbed by distracters. However, the accuracy of the CNFT model starts to decrease when the number of distracters added to the background is greater than 10 (see Figures 7(b)).

Again, the time that distracters remain unchanged modifies the accuracy of the model, see Figures 7(a) and 7(b). As for the noisy conditions, the accuracy is better with rapidly changing distracters (e.g. changing every computational step), but still highly acceptable when the positions of distracters are updated every two seconds, whereas the CNFT-based model accuracy decreases more rapidly with slightly more inertial distracters.

From these two first sets of experiments, we can state



(a)



(b)

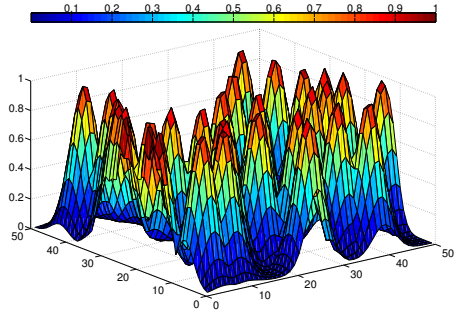
Fig. 4. A zero-mean Gaussian noise with different variances from 0 to 1 has been added to the input stimulus. The target changes its position every ten computational steps (one second) and the noise changes at differents rates (every 1, 5, 10, 15 or 20 computational steps of 0.1 second, as represented by the bars from left to right for each level of noise). (a) Accuracy of the proposed model in terms of the error level as a measure of the distance to the target using a focus map with $n = 50$ (using a normalized distance, i.e. computed in the $[0, 1] \times [0, 1]$ continuous field that is discretized onto the $n \times n$ neurons). (b) Accuracy of the CNFT model in terms of the error level as a measure of the distance to the target using a focus map with $n = 50$.

that the proposed model not only reproduces the robustness of the initial CNFT-based model, it clearly improves this robustness.

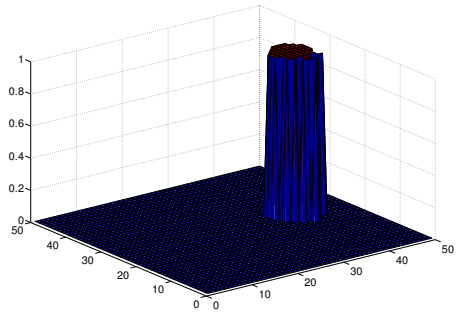
3) Robustness of the model with different values of γ: In this set of experiments, we modified the value of the γ factor in order to see its influence on the behavior of the proposed model.

All previous experiments were performed using $\gamma = 10$. In this section the same experiments are studied while setting γ with 1, 5, 15 and 20. Figure 8 shows the experimental results obtained using different values of γ: each bar shows the average accuracy of the model computed when the target and noise or distracters are updated at different rates.

The behavior of the model when the input is altered by noise and using different values of γ is shown in Figure 8(a).



(a)



(b)

Fig. 5. Distracters are exact copies of the input Gaussian patch, but they lack spatio-temporal continuity. (a) Input with 30 distracters (b) Neurons firing in the focus map.

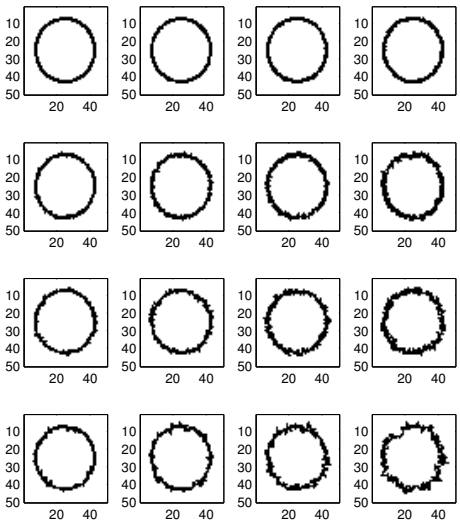
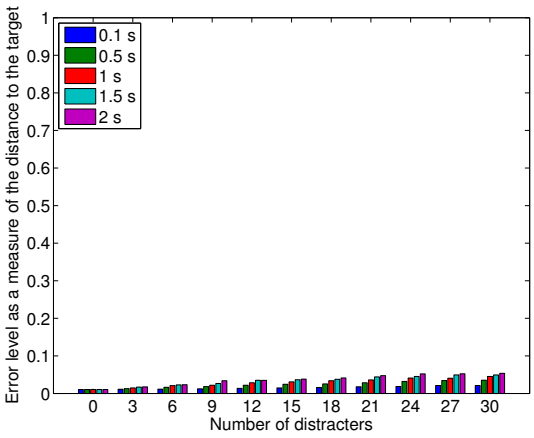
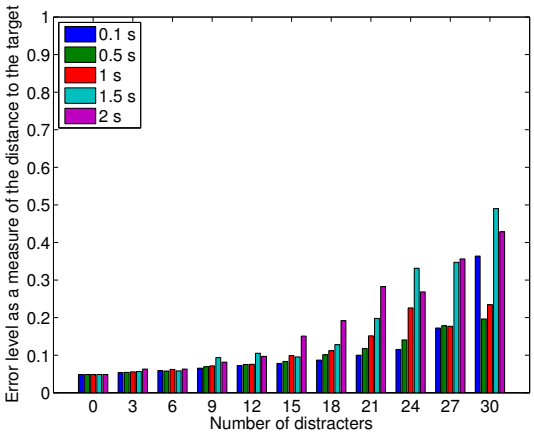


Fig. 6. Trajectories followed by the bubble of activity in the focus map when the target changes its position each ten computational steps. Each column presents the trajectory followed by the bubble when some distracters are added (3, 6, 12 and 24 distracters). Each row presents the trajectory followed by the bubble when distracters are updated at different rates (every 1, 5, 10 and 20 computational steps).

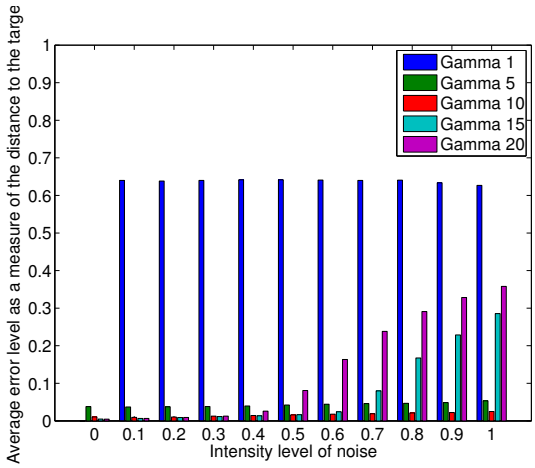


(a)

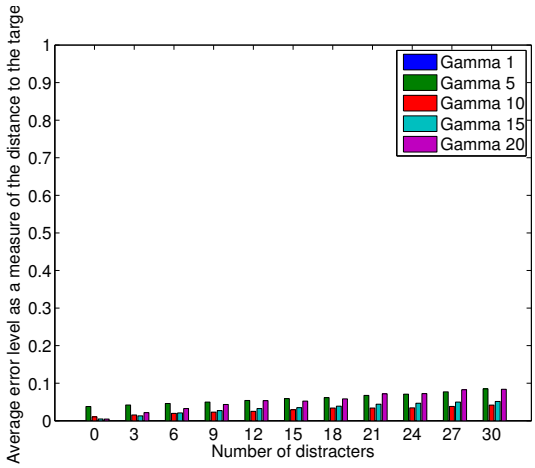


(b)

Fig. 7. Zero to 30 distracters (with same width and intensity as the original stimulus) were added to the stimulus. The target changes its position every ten computational steps (one second) and the distracters' positions change at different rates. (a) Accuracy of the proposed model in terms of the error level as a measure of the distance to the target using a focus map with $n = 50$. (b) Accuracy of the CNFT model in terms of the error level as a measure of the distance to the target using a focus map with $n = 50$.



(a)



(b)

Fig. 8. Average accuracy of the proposed model in terms of the error level as a measure of the distance to the target using a focus map with $n = 50$ and different values of γ (1, 5, 10, 15 and 20, as represented by the bars from left to right). The target changes its position every ten computational steps. (a) A zero-mean Gaussian noise with different variances from 0 to 1 is added to the input stimulus. The noise changes every ten computational steps. (b) Zero to 30 distracters (with same width and intensity as the original stimulus) are added to the stimulus. The distracters' positions change every ten computational steps.

The first fact that we have to point out is that when $\gamma = 1$ and the input stimulus lacks of noise, the neurons of the model do not spike at all; when some noise is added to the input stimulus the neurons of the model spike but the model is not able to focus on and track the input target. On the other hand, the accuracy of the model starts to decrease when the noise added to the input is greater that 0.6 and γ is greater than 15.

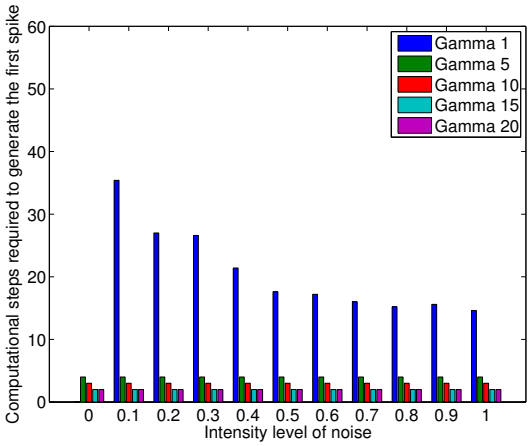
The behavior of the model when several distracters are added to the input with different values of γ is shown in Figure 8(b). In these experiments we observed that no matter the number of distracters, if $\gamma = 1$ the neurons of the model do not spike, whereas the behavior of the model is rather similar for all values of γ that are strictly greater than 1.

4) Reaction time of the proposed model: In this set of experiments we studied the influence of changing γ in terms of the reaction time of the proposed model. Using the same experimental setup described in the previous section, we count the number of computational steps that the model needs before the first spike emerges.

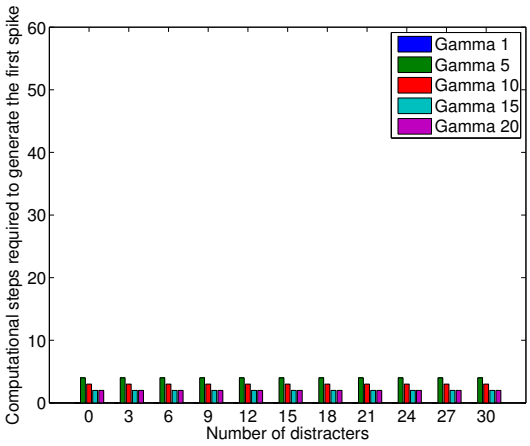
The number of computational steps required by the model in order to generate the first spikes are shown in Figure 9. When the value of γ is small, the model requires more computational steps to generate the first spikes in the focus map, see Figure 9(a). Nevertheless, this influence remains weak, except for the case $\gamma = 1$.

V. Conclusion

A simple visual attention model based of leaky integrate-and-fire neurons has been described in this paper. Through several experiments, this model has been proved to be very robust and able to track one moving target in the presence



(a)



(b)

Fig. 9. Computational steps required by the model in order to generate the first spike using different values of γ. The target changes its position every ten computational steps. (a) A zero-mean Gaussian noise with different variances from 0 to 1 is added to the input stimulus. The noise changes every ten computational steps. (b) Zero to 30 distracters (with same width and intensity as the original stimulus) are added to the stimulus. The distracters' positions change every ten computational steps.

of noise with very high intensity or in the presence of a lot of distracters.

The experimental results obtained with the proposal are comparable to those provided by previously defined CNFT-based and LIF-based models, in terms of both behavioral properties and robustness. However, the level of robustness is greatly improved with respect to the initial CNFT-based model, and, compared against the previous LIF-based model, no filtering technique is used and the computational resources reduce to only one neural map of LIF neurons.

Our model introduces a gaining factor that increases the influence of the input stimulus. The experimental study shows that its influence remains limited, except for its lowest values. Nevertheless, it could have a more significant influence when assembling several spiking neural maps to perform more complex visual tasks such as overt attention.

Though reducing the topological constraints of future hardware parallel implementations thanks to 1-bit communications between neurons (spikes), this model still is not able to be mapped onto hardware devices with a fully parallel approach because of the dense lateral connections. Current efforts aim at exploiting the robustness of the spike-based approach so as to reduce the range of the lateral connection kernel while maintaining highly satisfactory attentional properties.

REFERENCES

[1] L. Busse, K. C. Roberts, R. E. Crist, D. H. Weissman, and M. G. Woldorff, "The spread of attention across modalities and space in a multisensory object," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 51, pp. 18 751–18 756, 2005.

[2] J. H. Maunsell and S. Treue, "Feature-based attention in visual cortex." *Trends in neurosciences*, vol. 29, no. 6, pp. 317–322, 2006.

[3] W. James, *The Principles of Psychology, Vol. 1*. Dover Publications, 1950.

[4] L. G. Ungerleider and M. Mishkin, *Two Cortical Visual Systems*. MIT Press, 1982, ch. 18, pp. 549–586.

[5] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[6] L. Itti and C. Koch, "Computational modelling of visual attention." *Nature reviews. Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[7] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry." *Human neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[8] N. P. Rougier and J. Vitay, "Emergence of attention within a neural population," *Neural Netw.*, vol. 19, no. 5, pp. 573–581, 2006.

[9] J. Fix, J. Vitay, and N. Rougier, *Anticipatory Behavior in Adaptive Learning Systems: From Brains to Individual and Social Behavior*. Springer-Verlag Berlin Heidelberg, 2007, ch. A Computational Model of Spatial Memory Anticipation during Visual Search.

[10] J. Fix, N. Rougier, and F. Alexandre, "A dynamic neural field approach to the covert and overt deployment of spatial attention," *Cognitive Computation*, pp. 1–15, 2010.

[11] M. Cerda and B. Girau, "Bio-inspired visual sequences classification," in *Brain Inspired Cognitive Systems 2010 Brain Inspired Cognitive Systems 2010 - BICS 2010*, Madrid Spain, 2010, p. pp. 20.

[12] C. Torres-Huitzil, B. Girau, and C. Castellanos Snchez, "On-chip visual perception of motion : A bio-inspired connectionist model on FPGA," *Neural networks*, vol. 18, no. 5-6, pp. 557–565, 2005.

[13] B. Girau and C. Torres-Huitzil, "Massively distributed digital implementation of an integrate-and-fire LEGION network for visual scene segmentation." *Neurocomputing*, vol. 70, pp. 1186–1197, 2007.

[14] S. Chevallier and P. Tarroux, "Visual focus with spiking neurons," in *ESANN*, 2008, pp. 385–389.

[15] S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, no. 2, pp. 77–87, 1977.

[16] J. Taylor, "Neural bubble dynamics in two dimensions: Foundations," *Biological Cybernetics*, vol. 80, pp. 5167–5174, 1999.

[17] S. Chevallier and P. Tarroux, "Covert attention with a spiking neural network," in *International Conference on Computer Vision Systems (ICVS)*, ser. Lectures Notes in Computer Science, vol. 5008, 2008, pp. 56–65.

[18] B. Cessac, "A discrete time neural network model with spiking neurons. rigorous results on the spontaneous dynamics." *J. Math. Biol.*, vol. 56, no. 3, pp. 311–345, 2008.

[19] J.-C. Quinton, "Exploring and optimizing dynamic neural fields parameters using genetic algorithms," in *Proceedings of IEEE World Congress on Computational Intelligence (IJCNN 2010) (Barcelona, Spain)*, 2010.

[20] N. Rougier and A. Hutt, "Synchronous and asynchronous evaluation of dynamic neural fields," *Journal of Difference Equations and Applications*, 2011, to appear.